Comparison Wordclouds of Four Medical Conditions

George Fisher

Thursday, July 17, 2014

Summary

Whereas simple word clouds show the relative frequency of words tweeted about individual conditions, we have two other word clouds that compare and contrast the words used about the conditions.

1. **Comparison Word Cloud**
   This word cloud shows the words that are specific to each condition.

2. **Commonality Word Cloud**
   This word cloud shows the words that the tweets about the four conditions have in common
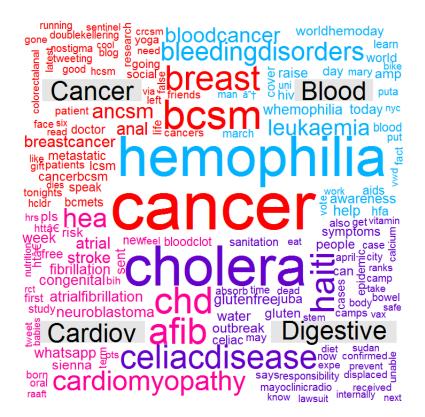
These are the four medical conditions we consider here:
- Blood Disorders
- Cancer
- Cardiovascular Diseases
- Digestive Dsorders

The corpus was restricted to the first 10,000 tweets in the database for each condition and then further reduced to include only those that had been retweeted more than three times.
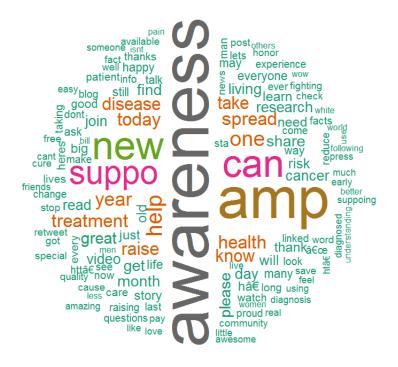
Comparison Word Cloud

Shows the words specific to the individual conditions.

Commonality Word Cloud

Shows the words that tweets about the four conditions have in common.

Credits and Description

```
# =======================================================================
# Title:        Comparison_Wordcloud.Rmd
# Author:       Gaston Sanchez
# Date:         May, 2012
#
# Modified by:  George Fisher
# Modified:     July 2014
#
# Description:  Simple script showing how to produce basics comparison
#               and commonality wordclouds in R from tweets of four
#               medical conditions.
#
# License:      BSD Simplified License
#               http://www.opensource.org/license/BSD-3-Clause
#               Copyright (c) 2012, Gaston Sanchez
#               All rights reserved
# =======================================================================
```

Programming Environment Details

```
## R version 3.1.0 (2014-04-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_1.0.0       wordcloud_2.5       RColorBrewer_1.0-5
## [4] tm_0.6              NLP_0.1-3
##
## loaded via a namespace (and not attached):
##  [1] colorspace_1.2-4 digest_0.6.4     evaluate_0.5.5   formatR_0.10
##  [5] grid_3.1.0       gtable_0.1.2     htmltools_0.2.4  knitr_1.6
##  [9] MASS_7.3-33      munsell_0.4.2    parallel_3.1.0   plyr_1.8.1
## [13] proto_0.3-10     Rcpp_0.11.2      reshape2_1.4     rmarkdown_0.2.46
## [17] scales_0.2.4     slam_0.1-32      stringr_0.6.2    tools_3.1.0
## [21] yaml_2.1.13
```